A Comparison among Support Vector Machine and other Machine Learning Classification Algorithms

Motaz M. H. Khorshid ¹, Tarek H. M. Abou-El-Enien ², Ghada M. A. Soliman ³

¹ Department of Operations Research & Decision Support, Faculty of Computers & Information, Cairo University,
5 Dr. Ahmed Zoweil St. - Orman - Postal Code 12613 - Giza – Egypt

² Department of Operations Research & Decision Support, Faculty of Computers & Information, Cairo University,
5 Dr. Ahmed Zoweil St. - Orman - Postal Code 12613 - Giza – Egypt

³ Department of Operations Research & Decision Support, Faculty of Computers & Information, Cairo University,
5 Dr. Ahmed Zoweil St. - Orman - Postal Code 12613 - Giza – Egypt

ABSTRACT
Predication of terrorist groups that responsible of terrorist attacks is a challenging task and a promising research area. There are many methods that have been developed to address this challenge ranging from supervised to unsupervised methods. The main objective of this research is to conduct a detailed comparative study among Support Vector Machine as one of the successful prediction classifiers that proved highly performance and other supervised machine learning classification and hybrid classification algorithms. Whereas most promising methods are based on support vector machine (SVM); so there is a need for a comprehensive analysis on prediction accuracy of supervised machine learning algorithms on different experimental conditions, and hence in this research we compare predictive accuracy and comprehensibility of explicit, implicit, and hybrid machine learning models and algorithms. This research based on predicting terrorist groups responsible of attacks in Middle East & North Africa from year 2004 up to 2008 by comparing various standard, ensemble, hybrid, and hybrid ensemble machine learning methods and focusing on SVM. The compared classifiers are categorized into main four types namely; Standard Classifiers, Hybrid Classifiers, Ensemble Classifiers, and Hybrid Ensemble Classifiers. In our study we conduct three different experiments on the used real data, afterwards we compare the obtained results according to four different performance measures. Experiments were carried out using real world data represented by Global terrorism Database (GTD) from National Consortium for the study of terrorism and Responses of Terrorism (START).

Keywords: Hybrid Models, Machine Learning, Predictive Accuracy, Supervised Learning.

1. INTRODUCTION
Machine learning (ML) is the process of estimating unknown dependencies or structures in a system using a limited number of observations [1]. ML algorithms are used in data mining applications to retrieve hidden information. Machine learning methods are rote learning, learning by being told, learning by analogy, and inductive learning, which includes methods of learning by examples and learning by experimentation and discovery [1][2]. Numerous machine learning methods and different knowledge representation models can be used for predicting different pattern in data set [3]. For example, classification, and regression methods can be used for learning decision trees, rules, Bayes networks, artificial neural networks and support vector machines. Supervised Machine learning classification is one of the tasks most frequently carried out by so called Intelligent Systems. Thus, a large number of techniques have been developed based on Artificial Intelligence (Logic-based techniques, Perceptron-based techniques) and Statistics (Bayesian Networks, Instance-based techniques). The concept of combing classifiers is proposed as a new direction for the improvement of the performance of individual machine learning algorithms. Hybrid and ensemble methods in machine learning have attracted a great attention of the scientific community over the last years [1]. Multiple, ensemble learning models have been theoretically and empirically shown to provide significantly better performance than single weak learners, especially while dealing with high dimensional, complex regression and classification problems [2]. Adaptive hybrid systems has become essential in computational intelligence and soft computing, a main reason for being popular is the high complementary of its components. The integration of the basic technologies into hybrid machine learning solutions [4] facilitate more intelligent search and reasoning methods that match various domain knowledge with empirical data to solve advanced and complex problems [5]. Both ensemble models and hybrid methods make use of the information fusion concept but in slightly different way. In case of ensemble classifiers, multiple but homogeneous, weak models are combined [6], typically at the level of their individual output, using various merging methods, which can be grouped into fixed (e.g., majority voting), and
trained combiners (e.g., decision templates) [7]. Support Vector Machine (SVM) is a type of learning algorithm developed in 1963; represents supervised machine learning approaches [8] and it is an excellent successful prediction method from examples. SVMs represent a new approach to pattern classification that has attracted a great deal of interest in the machine learning community. They operate on the induction principle of structural risk minimization, which minimizes an upper bound on the generalization error. SVMs have shown to be successful in solving many pattern recognition problems and perform much better than non-linear classifiers such as artificial networks in many situations [9]. The main concept of SVM is to obtain the Optimal Separating Hyperplane (OSH) between the positive and negative samples. This can be done through maximizing the margin between two parallel hyperplanes. Finding this plane, SVM can then forecast the classification of unlabeled sample through asking on which side of the separating plan the sample lies. SVMs are able to handle different types of classification problems such as linear and nonlinear classification problems. Both separable and non-separable problems are handled by SVMs in the linear and nonlinear case [10].

In our research paper, a real world data set of Middle East and North Africa is used for terrorism prediction based on comparative study among SVM prediction accuracy and the accuracy of multiple standard, hybrid, ensemble, and hybrid ensemble classification models and algorithms with the help of WEKA as an important and famous machine learning software written in JAVA [11], and hence, a detailed three experiments will be conducted to provide a detailed and clear comparison from the view of different four performance measures. The organization of this paper is as follows; Section 2 presents the literature review of the previous and present research in different types of ML classification algorithms. Section 3 illustrates the machine learning concept as well as machine supervised learning methods for learning classification. Section 4 explains hybrid machine learning models and methods. Section 5 studies in details the experimental methods, dataset and used software, results and analysis; illustrated with detailed diagram, and tables. Finally, section 6 covers conclusions and different future trends.

2. LITERATURE REVIEW

The foundations of Support Vector Machines (SVMs) have been developed by Vladimir Vapnik [12] and are gaining popularity due to many attractive features, and promising empirical performance. The formulation embodies the Structural Risk Minimization (SRM) principle, which has been shown to be superior, [13], to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks. SRM minimizes an upper bound on the VC dimension (‘generalisation error’), as opposed to ERM that minimizes the error on the training data, it is the difference which equips SVM with a greater ability to generalize, which is the goal in statistical learning. SVM were developed to solve the classification problem, but recently they have been extended to the domain of regression problems [14]. According to M. Wozniak [15] and J. Gama [16], there are a lot of machine hybrid methods developed in the past such as Model Trees-multivariate trees with linear or some other functional models at the leaves [17]–[18]–[19]. Perception Trees- combination of a decision tree and a linear threshold unit are presented by P. E. Utgoff [20]. B.Konoenko and R.Kohavi proposed a new algorithm, NBTtree, which induces a hybrid Decision Tree and Naïve Bayes Classifiers where the decision tree nodes contain univariate splits as regular decision-trees, but the leaves contain Naïve Bayes classifiers [21]–[22]. Functional trees- an extension of multivariate and model trees [12]. Model Class Selection- a hybrid algorithm that combines, in a single tree, nodes that are univariate tests, or multivariate tests generated by linear machines or instance-based learners [20]. In Meta Decision Tress- Lj. Todorovski, S. Dzeroski combined different classifiers with meta decision trees where leaves predict which classifier should be used to obtain a prediction [23]. Carla E Brodley and Paul E Utgoff presented stacked generalized hybrid ensembles which are constructed from different base learning methods [24]. In Hybrid Hoeffding Trees - several hybrid variants of the basic method using naive bays, functions and ensemble methods are presented by S. B. Kotsiantis and I. D. Zaharakis [25].

Piotr Sobolewski and Michal Wozniak [15] faced with concept drift that means the problem of significant changes in statistical properties of the target variables is usually caused by some hidden and unknown features making the classification models less accurate over course of time. Detection of concept drift is very important in real dynamic environments since it may be a hint to trigger classification model reconstruction, and they focus on detection of virtual concept drift using unsupervised learning based on knowledge about the possible data distributions that may occur in the data stream; without any knowledge about real class labels. A priori distribution patterns are treated as the known concepts, among which changes are being detected. The authors have developed their own method called simulated recurrence based on majority voting ensembles on results of statistical tests for distributions of known features. As an additional benefit, the concept detection makes the selection of the right classification model easier since a separate model may be pre-assigned to each concept. Javier Torres Niño et al. extended fundamental classification method – decision trees by combination unsupervised and supervised machine learning, i.e. clustering and classification. Additionally, they utilize a third component, which goal is to adjust clustering parameters. First, the predicted class attribute is removed before clustering and the number of instances in the majority class is calculated and compared with a given threshold to determine whether
the instances in the entire cluster are treated as classified or not. The instances from the non-classified cluster are used to learn the decision tree [26]. Tomasz Kajdanowicz and Przemysław Kazienko [27] provided a new method for the complex machine learning problem – multi-label classification, in which every instance can be independently assigned with many class labels simultaneously. The problem becomes especially demanding in case of larger output space – with many possible subsets of the class label set. The method is derived from the general boosting concept adapted to the multi-label environment. Chun-Wei Lin et al. [28] proposed a new integrated MFFP-tree algorithm to extract fuzzy association rules, its main feature is its ability to process and integrate multiple sources, local databases. It has been achieved by means of integration of many local fuzzy regions and tree branches into one coherent multiple fuzzy frequent pattern tree (MFFP-tree). It enables the authors to generate more complete global association rules, also preserving their local equivalences.

3. MACHINE LEARNING METHODS

3.1 Machine Learning

Machine learning (ML) is defined as the process of estimating unknown dependencies or structures in a system using a limited number of observations [1]. The goal of ML is to devise learning algorithms that do the learning automatically without human intervention or assistance. Machine learning tasks are classification, regression and clustering. Machine learning methods are rote learning, learning by being told, learning by analogy, and inductive learning, which include different methods of learning by examples, learning by experimentation, and discovery [11-12]. There are several applications for ML, the most significant of which is predictive data mining [25]. Every instance in any dataset used by ML algorithms is represented using the same set of features. The features may be continuous, categorical, or binary. If instances are given with known labels then learning is called supervised, in contrast with unsupervised learning, where instances are unlabeled [29]. Numerous ML applications involve tasks that can be set up as supervised. In our research, we have concentrated with machine learning of classifications which is an important branch of study. A system learns to classify new cases to predefined discrete problem classes. Classification is a special kind of regression, ML of classification performs an estimation of an unknown dependence between input (data) and output of the considered system (classification) [3].

The main goal of our research is to perform a comparative study among SVM as a successful prediction classifier and other different types of supervised ML classification algorithms and ensemble methods, and the possibility of combining classifiers which is usually better than any of its elements.

3.2 Machine Learning Methods for Learning Classifications

3.2.1 Methods for learning Comprehensible Knowledge

Methods for learning comprehensible, human readable knowledge are especially appropriate in building knowledge based decision support systems/expert systems. Well known and famous methods are Decision Tree (DT), and rule Learning (RL), as well as Hoeffding Tree or Very Fast Decision Tree (VFDT), which is a new method introduced for incremental machine learning from data streams [25]. It stores a data stream only once and after that updates the tree.

3.2.2 Methods for Learning Implicit Knowledge

Implicit or distributed knowledge is subjective, empirical, hard to formalize, and not understandable for humans [3]. It can be represented in forms of bayes or neural networks, support vectors or using the similarity function and learning examples by itself. The most used methods of this type are K-Nearest Neighbours (KNN), Bayes Networks, Artificial Neural Networks (ANN), and Support Vector Machine (SVM).

Support Vector Machines are a very specific class of algorithms, characterized by the use of kernels, the absence of local minima, the sparseness of the solution and the capacity control obtained by acting on the margin, or on other “dimension independent” quantities such as the number of support vectors.. SVMs is a new generation learning system based on recent advances in statistical learning theory. SVMs deliver state-of-the-art performance in real-world applications such as text categorization, hand-written character recognition, image classification, biosequences analysis, etc., and are now established as one of the standard tools for machine learning and data mining.

SVM is a very successful method of machine learning from examples [30], which is based on mapping of learning examples from input space to a new high dimensional, potentially infinite dimensional feature space in which examples are linearly separable. The method then finds an optimal hyperplane [31].

\[ \langle w, \phi(x) + b = 0 \rangle \]  

(1)

Where \( w \) is a matrix of coefficients \( \phi(x) \) is a mapping function, and \( b \) is a constant. This hypersurface separates learning examples with a maximal margin or distance to the nearest learning example [30]. Support vectors are a small set of critical border examples of each class, best separated by this hyperplane. Construction of an optimal hyperplane is performed using iterative algorithm which minimizes the error estimation function:
\[
\frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i
\]

with the constraints:

\[
y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \ldots, N, \quad \xi \geq 0, \quad i = 1, \ldots, n
\]

where \( w \) is a vector of coefficients, \( b \) is a constant, \( \xi \) is a slack variable (tolerance of overlapping linear non separable classes of examples), \( n \) is a number of learning examples and \( C \) is a regularization parameter. SVM method uses linear functions to create discrimination borders in a high dimensional space. Non-linear discriminant function in an input space is obtained using inverse transformation (kernel trick) [3].

### 3.2.3 Redundant Knowledge Machine Learning

Methods of learning and combining redundant classifiers or ensembles are one approach for increasing prediction accuracy models on unseen examples, which is the most important generalization property. The most famous methods in that regard are Random Forests Method [32], which simultaneously uses two sources of diversity of its elements: (1) resampling of learning data and (2) resampling the attribute set as part of the induction process, the other method called CART [33].

### 4. HYBRID MACHINE LEARNING METHOD

Supervised learning is the machine learning task of inferring a function from supervised training data [3]. This function is called a classifier; with other words, the supervised learning problem is to find an approximation to unknown function given a set of previously labeled examples. Different methods explore different hypothesis spaces, use different search strategies and are appropriate for different types of problems. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier (if the output is discrete, so we deal with classification) or a regression function (if the output is continuous, so it is a regression). The inferred function should predict the correct output value for any valid input object.

### 5. EXPERIMENTS

This research paper investigates and compares applicability of different types of supervised ML classifiers to predict the terrorist groups that responsible of attacks in Middle East & North Africa from year 2004 up to year 2008, and based on comparing the SVMs’ performance measures results among those different types of classification algorithms.

#### 5.1 Methods

Different types of classification algorithms are selected and compared namely; standard and hybrid machine learning together with ensemble and hybrid ensemble methods:

1) Standard Methods: Support Vector Machine (SVM), Naïve Bayes (NB), K-nearest neighbours (KNN), Decision Tree (C4.5/ J48), Bayes Net (BN), Simple Logistic (SL) which is a type of Artificial Intelligence Logic-based Technique.

2) Hybrid Methods: Hybrid Hoeffding Tree (HHT), Functional Tree (FT), Hybrid Naïve Bayes with Decision Table (DTNB), Classification via Clustering (C via C), LADTree (Hybrid Decision Tree using LogitBoost strategy).

3) Ensemble Methods: Random Forests (RFs), AdaBoost, LogitBoost, MultiBoost


#### 5.2 Data Set

The data set used in our research study is a real world data about terrorist events occurred in Middle East & North Africa in the period from 2004 till 2008, which consists from a total of 89973 terrorist events (instances), and 50 attributes, the attribute terrorist group is consisting from 110 diverse terrorist groups. Before applying different standard and hybrid classification algorithms usually some pre-processing is performed on the data set. There is a hierarchy of problems that are often encountered in data preparation and pre-processing [25]:

- Impossible values have been inputted.
- Unlikely values have been inputted.
- No values have been inputted (missing values).
- Irrelevant input features are present in the data at hand.
The main steps in our research paper are explained in the following flow diagram:

![Flow Diagram of Main Steps in the Research Study](image)

**Figure 1** Flow Diagram of Main Steps in the Research Study

In order to perform data processing, it is essential to improve the data quality [30]. There are a few number of techniques used for the purpose of data pre-processing as data aggregation, data sampling, dimension reduction, feature creation, data discretization, variable transformation, and dealing with missing values. It is necessary in our research to apply the following steps for data preparation and data pre-processing:

5.2.1 First Step
Data reduction is performed on the terrorism data by selecting the most informative attributes that are highly correlated to our predicted attribute (Terrorist Group Name) without lose any critical information for classification and so 10 attributes are selected to be included in our experiment. The selected attributes are year, month, country, region, provstate, city, attack-type, target-type, weapon-type, and group-name. These selected attributes are highly related to the predicted attribute (Terrorist Group).

Instance selection is not used to handle noise but to cope up with the infeasibility of learning from very large data sets; instance selection in our data set is an optimization problem that attempts to maintain the mining quality while minimizing the sample size [34]. It reduces data and enables a data mining algorithm to function and work effectively with very large data sets. There are a variety of procedures for sampling instances from a large data set. The most well known are [35] random sampling and stratified sampling.

In our data set, we applied stratified sampling as a supervised filter instance method which is applicable when the class values are not uniformly distributed in the training sets, instances of the minority class (es) are selected with greater frequency in order to even out the distribution.

5.2.2 Second Step
For the missing data values, there are three approaches to handle missing data elements: removal, imputation, and special coding [33], [36]. In our research, we applied the approach of Litwise Deletion or data removal for the unknown and
missing data instances in order to produce the new data, and then we will conduct our experiments by applying the selected classification algorithms on new data set and compare between them via the classification accuracy and other three different performance measures namely; precision, recall, and f-measure.

5.2.3 Third Step
Conducting the experiments to perform different classification algorithms on the research data set by using WEKA as one of important tools available for implementing data mining algorithms to train the base classifiers then the evaluation of the implemented classifiers is performed by using the testing data set.

5.3 Software used
The Machine learning classification algorithms in this research are implemented based on WEKA. Waikato Environment for Knowledge Analysis (WEKA) is open source software written in JAVA, a public collection of machine learning algorithms allows the researcher to mine his own data for trends and patterns. The algorithms can either be applied directly to a dataset or called from the researcher own JAVA code [11]. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. In our experiment we performed different supervised ML classification algorithms and models on the Terrorism data of Middle East & North Africa from 2004 to 2008, by using Litwise deletion of missing and unknown instances, and a stratified supervised instance reduction method with the help of WEKA Software. During experiment, and after pre-processing steps, the data file is converted to .ARFF file to be used under WEKA JAVA environment.

5.4 Experimental Methods
1. Experiment I
The experiment is conducted by using the whole data set as testing data set and the results are illustrated in TABLE 1.

2. Experiment II
The experiment is conducted based on percentage split of the whole data into 66% of the data for training the classification model and 34% of the data for testing the model as illustrated in TABLE 2.

3. Experiment III
The experiment is conducted by using 10 Fold cross validation as testing option and the results are shown in TABLE 3.

The results of the applied standard, hybrid, ensemble, and hybrid ensemble classifiers from the conducted experiments will be evaluated according to four performance measures which are defined bellow:
- The Classification Accuracy: is the percentage number of correctly classified instances (the number of correct predictions from all predictions made)
- Precision: is a measure of classifier exactness (used as a measure for the search effectiveness)
- Recall: is a measure of classifier completeness.
- F-Measure: also called F-Score, it conveys the balance between the precision and the recall.

5.5 Experiments’ Results and Analysis
5.5.1 Experiment I
TABLE 1 shows the classification accuracy and other performance measures results from conducting experiment I of applying various and different types of standard, hybrid, ensemble, and hybrid ensemble classifiers with focus on SVM and compare the results as follow:

In Standard Classification Algorithms; SVM, KNN, and SL classifiers are superior in their accuracy results where it could classify above 95% all the instances correctly, and so they have the highest measures. The researcher can notice too that C4.5 decision tree classifier has lowest accuracy since it classify correctly only 60% from the whole data.

In Hybrid Algorithms; it is obvious that LogitBoost hybrid classifier is superior as well as RFs algorithm in which they have highly accuracy than other classifiers, the researcher can notice that both Adaboost and MultiBoost classifiers are not accurate and performs badly in prediction and they have the same accuracy as well as the same performance results.

In Ensemble Method; Random Forests (RFs) method is very accurate where it classify correctly about 99% of the data; its precision, recall, and f-measure are high as well.

In hybrid Ensemble classifiers; we can notice that Stacking, StackingC, and Vote Hybrid ensemble classifiers have the same results and they have the worst accuracy and the lowest performance measures as they could not classify correctly more that 16.9% of the whole training data.

The overall comparison among SVM and other different classifiers showed that SVM is considered an accurate classifiers and outperformed many other supervised ML classification algorithms.
5.5.2 Experiment II

TABLE 2 shows the ML classification accuracy and other performance measures results from conducting experiment II in which the whole data set is divided into two splits; 66% for training the classifiers, and 34% for testing the classifier of applying the different types of classifiers and perform a comparative analysis between SVM and other classifiers as follow:

In Standard Classification Algorithms; the researcher can notice that SL classifier is fairly good as it could classify 50% from the 66% of the data correctly; SVM classified about 40% of the training. NB, KNN, and C4.5 are not accurate classifiers as they could not classify correctly more than 31% of the training data, and they have the same accuracy results as well as the same performance measures.

In Hybrid algorithms; LADTree and FT have the highest classification accuracy among the other hybrid classifiers, it is obvious that classification via clustering performs badly among the other in which it could not classify correctly more than 13% of the training split of the data.

It is obvious that HHT and classification via clustering hybrid classifiers are not accurate as they could not classify more than 32% and 37% from the training data as they also have bad precision and f-measure results.

In Ensemble Methods; LogitBoost classifiers performs well as it could classify 50% of the training data part. AdaBoost and MultiBoost ensemble classifiers have the same results.

In hybrid Ensemble classifiers; Stacking, stacking, and Vote classifiers have the same accuracy as well as the same precision, recall, and f-measure, it is noticeable that they perform badly and not accurate; as it could not classify correctly more than 20% from 66% of the training data.

The overall accuracy and performance measures in Table2 shows that SVM is not accurate in that type of experiment comparing with other classifiers, and standard classifiers outperformed the hybrid classification algorithm.

5.5.3 Experiment III

In this type of experiments the classifiers are testing according to 10 fold cross validation. The accuracy results and overall performance measures are explained in TABLE 3 as follow:

In Standards Classifiers Algorithms; SVM algorithms has highest accuracy as well as highest precision, recall, and f-measure in 10-fold cross validation experiment, then C4.5 decision tree performs fairly good. But KNN performs badly; it could not classify correctly over than 36% of the data.

In Hybrid Algorithms; it is obvious that FT and DTNB outperformed the other hybrid classifiers in their results and can be considered almost good classifiers. From other side LADTree performs bad in which it classify correctly only 16% from the data.

---

**TABLE 1: Performance Measures Results of Experiment I**

<table>
<thead>
<tr>
<th>Method</th>
<th>Correctly Classified (%)</th>
<th>Incorrectly Classified (%)</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>94.6154%</td>
<td>5.38460%</td>
<td>0.949</td>
<td>0.946</td>
<td>0.942</td>
</tr>
<tr>
<td>NB</td>
<td>76.1538%</td>
<td>23.8462%</td>
<td>0.713</td>
<td>0.762</td>
<td>0.712</td>
</tr>
<tr>
<td>KNN</td>
<td>99.2308%</td>
<td>0.76920%</td>
<td>0.993</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>C4.5</td>
<td>60%</td>
<td>40%</td>
<td>0.453</td>
<td>0.60</td>
<td>0.499</td>
</tr>
<tr>
<td>BN</td>
<td>73.8462%</td>
<td>26.1538%</td>
<td>0.708</td>
<td>0.738</td>
<td>0.686</td>
</tr>
<tr>
<td>SL</td>
<td>98.4615%</td>
<td>1.5385%</td>
<td>0.988</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>HHT</td>
<td>86.1538%</td>
<td>13.8462%</td>
<td>0.854</td>
<td>0.862</td>
<td>0.844</td>
</tr>
<tr>
<td>FT</td>
<td>98.4615%</td>
<td>1.5385%</td>
<td>0.986</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>DTNB</td>
<td>53.0769%</td>
<td>46.9231%</td>
<td>0.38</td>
<td>0.531</td>
<td>0.421</td>
</tr>
<tr>
<td>C via C</td>
<td>20.7921%</td>
<td>79.2079%</td>
<td>0.043</td>
<td>0.208</td>
<td>0.072</td>
</tr>
<tr>
<td>LADT.</td>
<td>69.2308%</td>
<td>30.7692%</td>
<td>0.630</td>
<td>0.692</td>
<td>0.646</td>
</tr>
<tr>
<td>RFs</td>
<td>98.4615%</td>
<td>1.5385%</td>
<td>0.987</td>
<td>0.985</td>
<td>0.985</td>
</tr>
<tr>
<td>AdaB.</td>
<td>29.2308%</td>
<td>70.7692%</td>
<td>0.104</td>
<td>0.292</td>
<td>0.149</td>
</tr>
<tr>
<td>LogitB.</td>
<td>99.2308%</td>
<td>0.7692%</td>
<td>0.993</td>
<td>0.992</td>
<td>0.992</td>
</tr>
<tr>
<td>MultiB.</td>
<td>29.2308%</td>
<td>70.7692%</td>
<td>0.104</td>
<td>0.292</td>
<td>0.149</td>
</tr>
<tr>
<td>Stacking</td>
<td>16.9231%</td>
<td>83.0769%</td>
<td>0.029</td>
<td>0.169</td>
<td>0.049</td>
</tr>
<tr>
<td>Stack.C.</td>
<td>16.9231%</td>
<td>83.0769%</td>
<td>0.029</td>
<td>0.169</td>
<td>0.049</td>
</tr>
<tr>
<td>Vote</td>
<td>16.9231%</td>
<td>83.0769%</td>
<td>0.029</td>
<td>0.169</td>
<td>0.049</td>
</tr>
</tbody>
</table>
**TABLE 2: Performance Measures Results of Experiment II**

<table>
<thead>
<tr>
<th>Method</th>
<th>Correctly Classified</th>
<th>Incorrectly Classified</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>46.1538%</td>
<td>53.8462%</td>
<td>0.389</td>
<td>0.462</td>
<td>0.413</td>
</tr>
<tr>
<td>NB</td>
<td>42.3077%</td>
<td>57.6923%</td>
<td>0.343</td>
<td>0.423</td>
<td>0.343</td>
</tr>
<tr>
<td>KNN</td>
<td>36.1538%</td>
<td>63.8462%</td>
<td>0.313</td>
<td>0.362</td>
<td>0.329</td>
</tr>
<tr>
<td>C4.5</td>
<td>45.3846%</td>
<td>54.6154%</td>
<td>0.286</td>
<td>0.454</td>
<td>0.346</td>
</tr>
<tr>
<td>BN</td>
<td>41.5385%</td>
<td>58.4615%</td>
<td>0.268</td>
<td>0.415</td>
<td>0.313</td>
</tr>
<tr>
<td>SL</td>
<td>40%</td>
<td>60%</td>
<td>0.314</td>
<td>0.40</td>
<td>0.339</td>
</tr>
<tr>
<td>HIT</td>
<td>20.7692%</td>
<td>79.2308%</td>
<td>0.106</td>
<td>0.208</td>
<td>0.117</td>
</tr>
<tr>
<td>FT</td>
<td>42.3077%</td>
<td>57.6923%</td>
<td>0.369</td>
<td>0.423</td>
<td>0.391</td>
</tr>
<tr>
<td>DTNB</td>
<td>41.5385%</td>
<td>58.4615%</td>
<td>0.231</td>
<td>0.415</td>
<td>0.288</td>
</tr>
<tr>
<td>C via C</td>
<td>22.3077%</td>
<td>77.6923%</td>
<td>0.083</td>
<td>0.223</td>
<td>0.120</td>
</tr>
<tr>
<td>LADT.</td>
<td>40.9091%</td>
<td>59.0909%</td>
<td>0.345</td>
<td>0.409</td>
<td>0.345</td>
</tr>
<tr>
<td>RFs</td>
<td>35.3846%</td>
<td>64.6154%</td>
<td>0.293</td>
<td>0.354</td>
<td>0.306</td>
</tr>
<tr>
<td>AdaB.</td>
<td>29.2308%</td>
<td>70.7692%</td>
<td>0.104</td>
<td>0.292</td>
<td>0.149</td>
</tr>
<tr>
<td>LogitB.</td>
<td>47.6923%</td>
<td>52.3077%</td>
<td>0.406</td>
<td>0.477</td>
<td>0.434</td>
</tr>
<tr>
<td>MultiB.</td>
<td>29.2308%</td>
<td>70.7692%</td>
<td>0.104</td>
<td>0.292</td>
<td>0.149</td>
</tr>
<tr>
<td>Stacking</td>
<td>40.9091%</td>
<td>59.0909%</td>
<td>0.345</td>
<td>0.409</td>
<td>0.345</td>
</tr>
<tr>
<td>Stack.C.</td>
<td>36.3636%</td>
<td>63.6364%</td>
<td>0.333</td>
<td>0.364</td>
<td>0.301</td>
</tr>
<tr>
<td>Vote</td>
<td>16.9231%</td>
<td>83.0769%</td>
<td>0.029</td>
<td>0.169</td>
<td>0.049</td>
</tr>
</tbody>
</table>

In Ensemble Methods; the researcher can notice that LogitBoost ensemble classifier performs fairly good as it classify about 48% correctly. The classifiers AdaBoost and MultiBoost have the same results.

In Hybrid Ensemble Classifiers; we can see that Stacking and Vote classifiers have the same performance results, but Sacking C performs badly in which it could not classify correctly than 14% of the data.

The overall results showed the superiority of SVM over all other classifiers types especially in 10 fold - cross validation experiment. It is obvious that hybrid ensemble classifiers have the same worst performance results.

**TABLE 3: Performance Measures Results of Experiment III**

<table>
<thead>
<tr>
<th>Method</th>
<th>Correctly Classified</th>
<th>Incorrectly Classified</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>40.9091%</td>
<td>59.0909%</td>
<td>0.345</td>
<td>0.409</td>
<td>0.345</td>
</tr>
<tr>
<td>NB</td>
<td>31.8182%</td>
<td>68.1818%</td>
<td>0.213</td>
<td>0.318</td>
<td>0.252</td>
</tr>
<tr>
<td>KNN</td>
<td>31.8182%</td>
<td>68.1818%</td>
<td>0.430</td>
<td>0.318</td>
<td>0.349</td>
</tr>
<tr>
<td>C4.5</td>
<td>31.8182%</td>
<td>68.1818%</td>
<td>0.334</td>
<td>0.318</td>
<td>0.289</td>
</tr>
<tr>
<td>BN</td>
<td>36.3636%</td>
<td>63.6364%</td>
<td>0.333</td>
<td>0.364</td>
<td>0.301</td>
</tr>
<tr>
<td>SL</td>
<td>50%</td>
<td>50%</td>
<td>0.366</td>
<td>0.500</td>
<td>0.408</td>
</tr>
<tr>
<td>HIT</td>
<td>20.4545%</td>
<td>79.5455%</td>
<td>0.042</td>
<td>0.205</td>
<td>0.069</td>
</tr>
<tr>
<td>FT</td>
<td>43.1818%</td>
<td>56.8182%</td>
<td>0.433</td>
<td>0.432</td>
<td>0.412</td>
</tr>
<tr>
<td>DTNB</td>
<td>34.0909%</td>
<td>65.9091%</td>
<td>0.208</td>
<td>0.341</td>
<td>0.254</td>
</tr>
<tr>
<td>C via C</td>
<td>15.9091%</td>
<td>84.0909%</td>
<td>0.051</td>
<td>0.159</td>
<td>0.072</td>
</tr>
<tr>
<td>LADT.</td>
<td>45.4545%</td>
<td>54.5455%</td>
<td>0.440</td>
<td>0.455</td>
<td>0.443</td>
</tr>
<tr>
<td>RFs</td>
<td>27.2727%</td>
<td>72.7273%</td>
<td>0.239</td>
<td>0.273</td>
<td>0.224</td>
</tr>
<tr>
<td>AdaB.</td>
<td>36.3636%</td>
<td>63.6364%</td>
<td>0.150</td>
<td>0.364</td>
<td>0.207</td>
</tr>
<tr>
<td>LogitB.</td>
<td>50%</td>
<td>50%</td>
<td>0.557</td>
<td>0.500</td>
<td>0.510</td>
</tr>
<tr>
<td>MultiB.</td>
<td>36.3636%</td>
<td>63.6364%</td>
<td>0.150</td>
<td>0.364</td>
<td>0.207</td>
</tr>
<tr>
<td>Stacking</td>
<td>20.4545%</td>
<td>79.5455%</td>
<td>0.042</td>
<td>0.205</td>
<td>0.069</td>
</tr>
<tr>
<td>Stack.C.</td>
<td>20.4545%</td>
<td>79.5455%</td>
<td>0.042</td>
<td>0.205</td>
<td>0.069</td>
</tr>
</tbody>
</table>

**6. CONCLUSION AND FUTURE TRENDS**

Different types of supervised standard, ensemble, and hybrid machine learning classification algorithms and models are introduced in this research paper with the main focus on SVM classifier for prediction of the terrorist groups responsible of...
terrorist attacks in Middle East and North Africa from year 2004 up to 2008, by conducting different three experiments, the data used in our experimental study is based on real world data represented by Global terrorism Database (GTD) from National Consortium for the study of terrorism and Responses of Terrorism (START).

To achieve the goal of this research; three different experiments are conducted on the used data, as well as using Litwise deletion approach to handle the missing data and provide a detailed comparative study and clear analysis of the used 18 classification algorithms which categorize into four main types namely; standard classification algorithms, hybrid classifiers, ensemble classifiers, and ensemble hybrid classifier. We based on using popular software in that area of study called WEKA software which based on JAVA environment and evaluate the obtained results via three different test options which are: evaluation on training set, split of the whole data into 66% for training the models and 34% for testing the model, and the third option is 10 fold cross-validation.

The overall results of the first experiment that conducted using the whole training data showed that SVM classification algorithm is considered an accurate classifiers and outperformed many other supervised ML standard, hybrid, ensemble, and hybrid ensemble classification algorithms. Whereas in the second experiment which divides the whole data into 66% for training the classifiers, and 34% for testing and validation of the classifiers; the overall results and performance measures proved that SVM is not accurate in that type of experiment comparing with other classifiers, and standard classifiers outperformed the hybrid classification algorithm. And in the last experiment which test the classifier using 10 fold cross validation showed the superiority of SVM among the other types of ML classification algorithms and models.

From the other side; we can conclude that hybrid machine learning classifiers demonstrate good and proved obvious improvement in predictive accuracy over some standard comprehensible and ensemble methods.

And so the overall performance of the different types of classifiers used proved that hybrid machine learning classifiers perform accurate and in some situations it could outperformed the single classifiers with some enhancement, but ensemble methods are more accurate and outperformed the hybrid ensemble methods in their prediction of terrorist groups’ attacks results.

**FUTURE TRENDS**

For Future research, there is a plan to make further hybridization between Support Vector Machines algorithms with metaheuristic and soft computing artificial intelligence algorithms to enhance SVMs parameters and improve the performance of that type of successful famous ML algorithms. As well as make a hybridization between different ML classification algorithms with Genetic Algorithms, and Neural Networks to improve the performance of hybrid classifiers. Other researchers could try to make the hybridization between GA and SVM to improve the classification accuracy and convergence speed.

Some researchers may perform a modification for this research by using different approaches for handling missing data instances, and feature selection then perform a comparative study. Other researchers may use different test options to test the performance of the classification algorithms.

**References**


AUTHORS

Prof. Dr. Motaz M. H. Khorshid, is a Professor of Operations Research, - Department of O.R. & D. S. - Faculty of Computers & Information-Cairo University. His research interests in decision analysis, soft computing, modeling, simulation and data mining

Dr. Tarek H. M. Abou-El-Enien is an Associate Professor of Mathematical Programming & Operations Research, - Department of O.R. & D. S. - Faculty of Computers & Information- Cairo University. His research interests in decision analysis, soft computing, modeling, data mining and multi-level multi-objective Programming field

Ghada M. A. Soliman received her degree in B. of science in Operations Research and Decision Support, at the faculty of Computers and Information, Cairo University in 2001. Then she received her Master Degree in Operations Research in 2007, she is currently a PhD student. Her employment experience includes FCI institution since 2001; she is employed with the Department of Operations Research as Lecturer Assistant. Her research interests include modeling, simulation, data mining and soft computing.